基于Catboost的AMSR-2半经验地表温度降尺度

李永康[1,2]， 王新军[1]， 马燕飞[3]， 胡贵锋[1]， 桂海月[1]， 张冠宏[1]

（1. 新疆农业大学草业与环境科学学院,新疆土壤与植物生态过程实验室,新疆 乌鲁木齐 830052；
2. 中国科学院空天信息创新研究院,数字地球重点实验室,北京 100094；
3. 邯郸学院经济管理学院地理系,河北 邯郸 056005）

摘 要：以古尔班通古特沙漠为研究对象,采用2019年AMSR-2的4通道被动微波亮温及MODIS植被指数,探索Catboost算法对被动微波地表温度空间降尺度的可行性,填补古尔班通古特沙漠MYD11A1的缺失像元,为获取全天候日、夜间多层土壤温度提供数据参考。结果表明:(1) 古尔班通古特沙漠日、夜间用于降尺度研究的特征向量(23.8 GHz V、36.5 GHz V、18.7 GHz H、89 GHz V、36.5-23.8 GHz V、36.5 V-18.7 GHz H、EVI、NDVI)与地表温度的相关性空间分异特征明显,显示沙漠相关性高,绿洲相关性低,且日间分异性更强;而盐矿覆盖会降低微波与地表温度相关性。(2) Catboost 4通道模型建立的被动微波亮温与地表温度间映射关系表现出鲁棒性。降尺度结果精度较高,日夜间R^2分别为0.977、0.980,RMSE分别为3.69 K、2.38 K,MAE分别为2.71 K、1.70 K。(3) 单通道相关性统计结果与重要性分析结果存在差异,表明特征因子相关性统计结果不能直接作为Catboost被动微波反演地表温度特征选择依据。(4) 降尺度结果与站点6层土壤温度均呈极显著正相关,并随着深度的增加相关系数r总体呈下降趋势,RMSE呈增加趋势。

关键词：地表温度；AMSR-2；Catboost；降尺度；古尔班通古特沙漠

地表温度(Land Surface Temperature, LST)作为长波辐射与地气湍流热通量交换的直接驱动力[1],是反映区域乃至全球尺度水热平衡的重要参数之一[2],及时有效地获取区域地表温度信息对气候、农业研究具有重要的现实意义[3-4],地表温度已广泛应用于城市热岛效应[5]、干旱监测[6]、能量平衡评估[7]、蒸散发[3-4]等研究领域。由于地形、土壤、植被、天气等因素的复杂性,造成地表温度空间上具有异质性[8],时间上具有波动性,需要高时空分辨率的数据描述地表温度的空间分布及时间演变。目前,包括MODIS与AVHRR在内的热红外及其他光学遥感晴空地表温度反演算法已达较高精度[9-10],但非晴空区域反演受天气影响严重,如中纬度地区MODIS数据超过60%被天气影响,云为最主要原因[11]。微波通道能穿透云层,受大气影响较弱,可获取复杂天气状况下的地表辐射信息[12-13],但其空间分辨率较低,在区域尺度研究中需融合其他数据来提高空间分辨率。微波地表温度空间降尺度是指在确定低空间分辨率微波像元值的基础上,融合高空间分辨率地表温度信息,使得微波全天候观测地表温度的同时,获取更多的细节信息[14]。

辐射传输理论是多通道被动微波反演地表温度的物理基础,但为病态反演,且部分参数机理有待明确[15-17]。基于物理过程的半经验反演算法对被动微波反演地表温度具有重要意义。最早,McFarland等[15]以SSM/I(Special Sensor Microwave/Imager)为数据源,通过多元线性回归对地表温度反演,发现37 GHz与地表温度相关性最高,其他通道处理可用于修正水汽影响。Njoku[18]、Gao[17]等以AMSR-E为数据源,表明36.5 GHz垂直极化为地表温度最佳反演通道,受到大气影响较小,能够获取浅层温度信息,但裸地的低含水量及散射作用会降低微波反演的精度。Mao[19-20]等以AMSR-E为数据源,建立单通道亮温与MODIS地表温度之间的关系,得出89

收稿日期：2020-12-10； **修订日期：**2021-05-05
基金项目：国家自然科学基金项目(41761085,41301205);自治区研究生科研创新项目(XJ2020G152)
作者简介：李永康(1995-),男,硕士研究生,主要从事陆面数据同化研究. E-mail: yongkang_xau@163.com
通讯作者：王新军. E-mail: wxj8112@163.com

GHz垂直极化相关性最高，36 GHz垂直极化与18 GHz水平极化和23 GHz垂直极化的差分别能够消除地表和大气水分对地表温度反演的影响，在此基础上建立经典的4通道算法。近年机器学习模型为解决此问题提供了新的思路，Sun[21]、Tan[22]等分别使用回归树与卷积神经网络融合AMSR-2多通道亮温对美国和中国全天候地表温度进行了反演，取得了较好的结果。机器学习主要是基于数值关系算法，能充分利用被动微波亮温的多通道信息，达到较高的反演精度，但单纯的机器学习方式，缺乏对通道的选择及其物理意义的分析，结果的无法解释性可能限制其进一步应用。新的机器学习算法往往能够推动半经验反演模型的发展。Catboost（Categorical boosting）是Yandex于2017年提出的一种新的GBDT（Gradient Boosting Decision Tree）鲁棒性算法，精度与稳定性均高于随机森林与支持向量机[23]，能够改善以往GBDT算法（如LightGBM、XGboost）的过度拟合问题，已逐渐应用于蒸散发反演[23-24]、土地分类[25]等研究。目前，基于Catboost算法的遥感研究相对较少，因此本研究结合Catboost算法与4通道组合，探索基于半经验模型的被动微波地表温度降尺度方法。

古尔班通古特沙漠区站点稀少，地表温度数据稀缺，仅有位于南缘绿洲区的新疆阜康荒漠生态系统国家野外科学观测研究站（简称阜康荒漠生态站），难以满足模型对输入数据的要求；且沙漠区域遥感地表温度产品受云影响严重，2019年MODIS地表温度产品日夜间云覆盖率分别达63.0%、65.9%，时间、空间上的不连续，严重影响了数据的使用及进一步研究。本文以古尔班通古特沙漠为研究区，通过Catboost算法建立基于空间结构相似性的精细图像与粗图像之间的非线性映射关系，将AMSR-2数据融入到地表温度降尺度研究中，为古尔班通古特沙漠提供全天候高空间分辨率的地表温度产品，拓宽Catboost算法应用领域，为数据稀缺区地表温度获取提供有效的方法。

1 数据与方法

1.1 研究区概况

古尔班通古特沙漠（44°15′~46°5′N，84°50′~91°20′E）是中国第二大沙漠（图1），位于准噶尔盆地中部，玛纳斯河流域以东及乌伦古河以南地区，面积约4.88×10^4 km^2，主要沙丘类型为沙垅，占沙漠面积80%[26]。属典型中温带大陆性干旱气候，常年受西风带控制，受此影响纵向沙丘从西北向东南移动。全年平均气温5~5.7 ℃，昼夜温差10~30 ℃，极端气温最高可达41.7 ℃，最低可至-31 ℃[27]。全年蒸发量为2000~2800 mm，降水主要集中于5—9月，降水量＜220 mm，沙漠腹地低至70~100 mm，冬季积雪深度约30 cm[28]，受此影响春夏季短命植物生长旺盛。

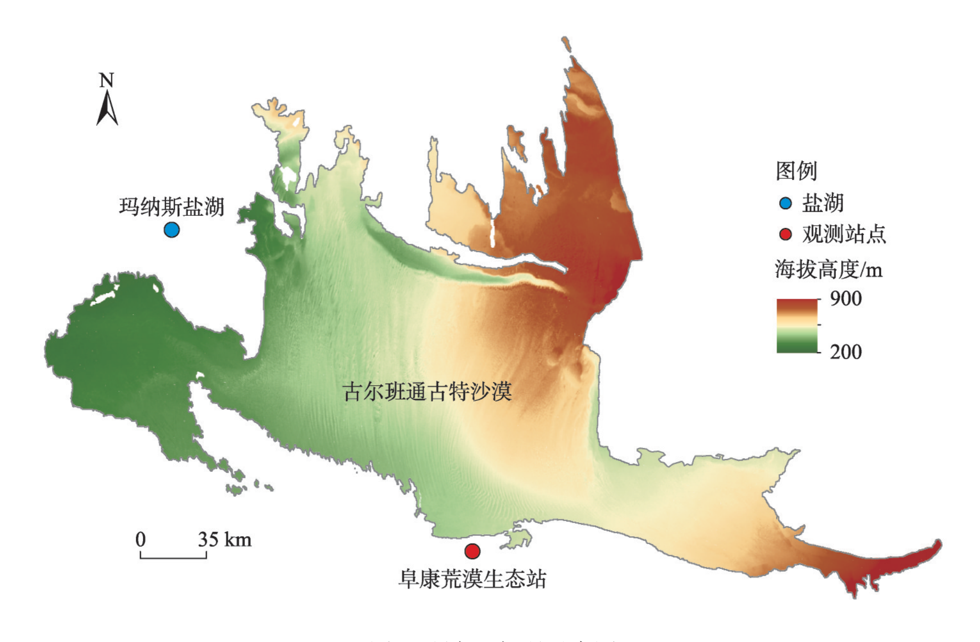


图1 研究区概况示意图

Fig. 1 Overview of study area

1.2 数据来源与处理

1.2.1 遥感数据　选取2019年GCOM-W卫星AMSR-2传感器的L1R数据集与Aqua卫星MODIS传感器的MYD13A3数据集作为Catboost模型特征向量，MYD11A1数据作为模型训练的目标向量(表1)。

L1R数据集由JAXA(http://gportal.jaxa.jp)提供，共包括6.9 GHz、10.7 GHz、18.7 GHz、23.8 GHz、36.5 GHz和89.0 GHz频率垂直与水平极化的14个通道。根据Mao等[19]4通道算法，选取36.5 GHz V亮温为温度反演的主导因子；23.8 GHz V与18.7 GHz H的亮温作为辅助数据，以消除大气、地表水等影响；89 GHz V消除其他因素的平均影响。单个扫描周期(16 d)共14 d过境研究区，过境时间约在日间15:00、夜间4:00。MYD13A2与MYD11A1数据由LADDC提供。MYD13A2为16 d合成增强植被指数(Enhanced Vegetation Index, EVI)与归一化植被指数(Normalized Vegetation Index, NDVI)，MYD11A1为日尺度地表温度数据。

AMSR-2被动微波亮温与MODIS-Aqua数据集获取时间相近，研究区二者单个扫描周期内每天过境时间差＜15 min，可减少因观测时间错位引起的误差；MYD11A1由于云影响，像元缺失严重，因此选用未被云影响且质量控制average LST error≤1 K的MYD11A1像元与AMSR-2被动微波数据集建立映射关系。

1.2.2 地表实测数据　阜康荒漠生态站(87°33′36″E、44°10′30″N)位于古尔班通古特沙漠南缘，是绿洲荒漠过渡区，地面植被较为茂盛，多以荒漠草原和草甸草原为主[29]，平均海拔461 m，提供每日逐时6层地表温度数据(分别距地表5 cm、10 cm、20 cm、40 cm、80 cm、160 cm)。由于AMSR-2扫描间隙的存在及自动站采集探头不明原因导致的数据缺失，共选白天264条、夜间270条数据参与地表温度验证。自动站数据由国家生态科学中心提供(http://rs.cern.ac.cn/index.jsp)。

1.2.3 数据处理　选取被动微波36.5 GHz V、23.8 GHz V、18.7 GHz H、89 GHz V亮温通道为特征向量进行组合变换，MYD13A2植被指数作为辅助数据，构建基于Catboost的半经验被动微波降尺度模型，将10 km分辨率被动微波亮温降尺度为1 km分辨率地表温度数据。AMSR-2日间与夜间地表温度降尺度具体步骤如下：

① 将AMSR-2与MODIS数据通过最近邻法统一重采样至10 km，选取MYD11A1质量average LST error≤1 K的像元匹配AMSR-2微波亮温数据及MYD13A2植被指数数据。

② 降尺度特征向量的空间相关性统计方法。在10 km空间分辨率下，6个AMSR-2被动微波亮温通道(4个原始单通道和2个组合通道)及2个MYD13A3植被指数与MYD11A1，在全年尺度(365 d)上进行单个像元位置的Pearson相关分析，获取各个特征向量的相关系数r及概率P数据，绘制相关系数r空间分布图(图2)；对研究区内相关系数r进行统计分析，绘制相关系数r频率分布图(图3)。进而分析各特征因子与地表温度的相关性及相关性的空间特征。

③ Ctboost 4通道模型十折交叉验证。基于前人研究成果及相关性空间统计分析，选取4通道半经验向量组合[21](36.5 GHz V、36.5−23.8 GHz V、36.5 V−18.7 GHz H、89 GHz V)结合MYD13A3增强植被指数训练Catboost模型，构建各特征因子与地表温度间的映射关系。通过“十折交叉验证”方案[30-31](10-fold cross-validation)评估模型。

④ 基于空间相似性的地表温度降尺度。将被动微波亮温通过最近邻算法重采样至1 km，与对应像元的1 km MYD13A3植被指数组成特征向量，输

表1　AMSR-2与MODIS相关数据参数

Tab. 1　The information of AMSR-2 microwave brightness temperature and MODIS data

AMSR-2被动微波亮温数据			MODIS数据		
中心频率/GHz	极化方式	空间分辨率/km	数据类型	空间分辨率/km	数据集
18.7	V/H	10	MYD 11A1	1	LST_Day_1 km
23.8	V/H	10			QC_Day
36.5	V/H	10	MYD 13A2	1	1 km_16_days_EVI
89	V/H	5			1 km_16_days_NDVI

注：H代表水平极化，V代表垂直极化。

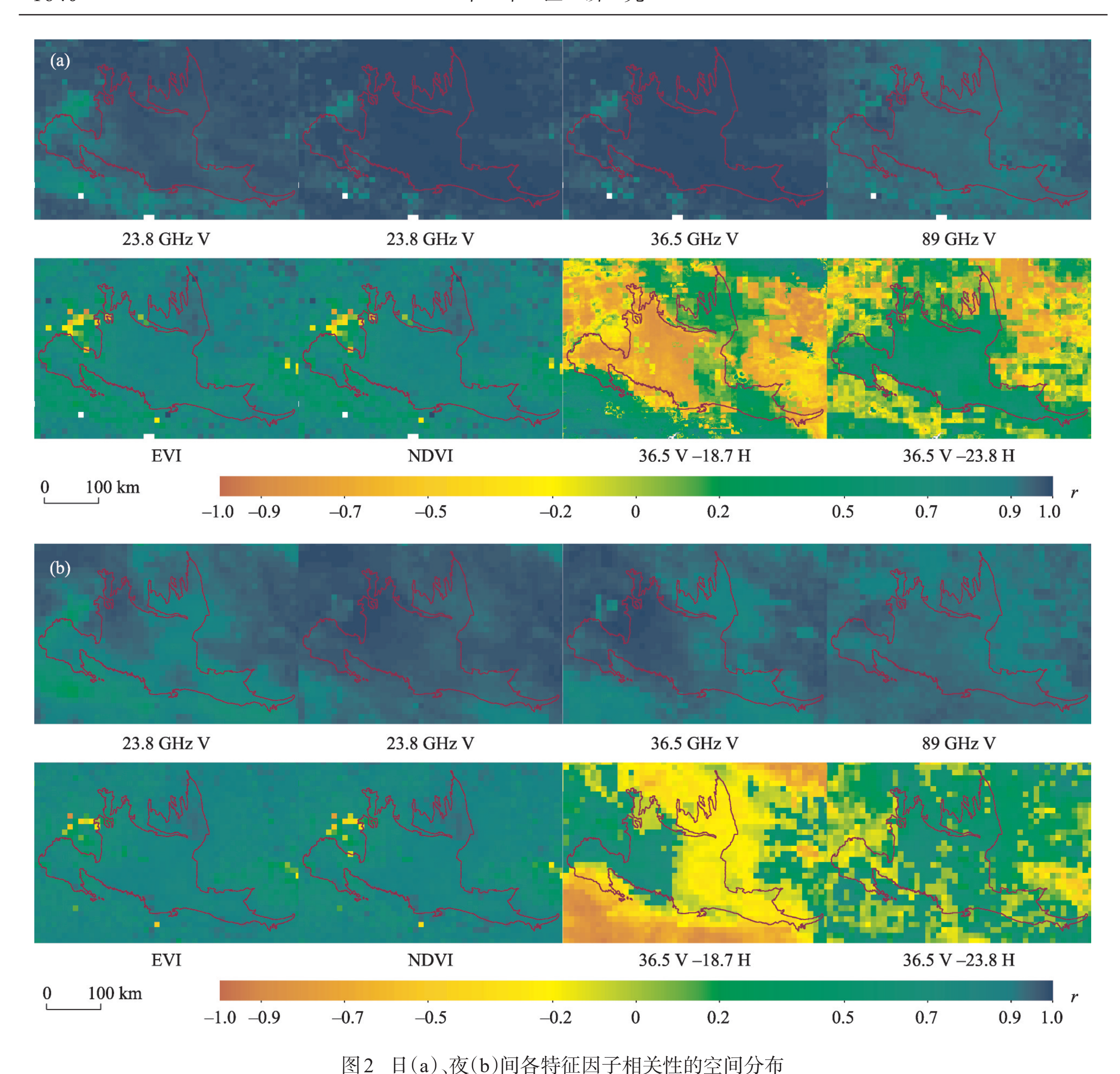


图2 日(a)、夜(b)间各特征因子相关性的空间分布

Fig. 2 Factors' correlations spatial distribution maps in the daytime (a) and nighttime (b)

入至10 km空间尺度上训练完成的Catboost模型,基于不同尺度空间结构相似性反演1 km空间分辨率的地表温度数据。

⑤ 降尺度结果的精度验证及特征因子的重要性评价。选用1 km空间分辨率质量控制average LST error ≤1 K的MODIS晴空像元,对降尺度结果进行精度验证;对模型因子进行重要性分析。

⑥ 降尺度结果的多层地温相关性分析。通过阜康荒漠生态站6层地温数据与对应位置的降尺度结果进行相关性分析,通过r、P、RMSE(Root Mean Square Error)对相关性结果进行评价。

⑦ 将基于Catboost半经验模型降尺度的被动微波地表温度与MODIS质量控制后的晴空像元融合,填补MODIS地表温度缺失像元,得到全天候的地表温度产品。

基于AMSR-2被动微波亮度温度的全天候地表温度降尺度技术路线如图4。

1.3 研究方法

1.3.1 *微波地表温度反演的辐射传输模型* 被动微波观测反演地表温度的物理基础为辐射传输理论[18],通过辐射能量平衡建立地表温度反演算法。辐射传输方程描述了被动微波辐射计接收到的总能量,包括地面辐射和向上、向下的大气路径辐射以及由大气吸收而减弱的辐射分量[19-20],通用辐射

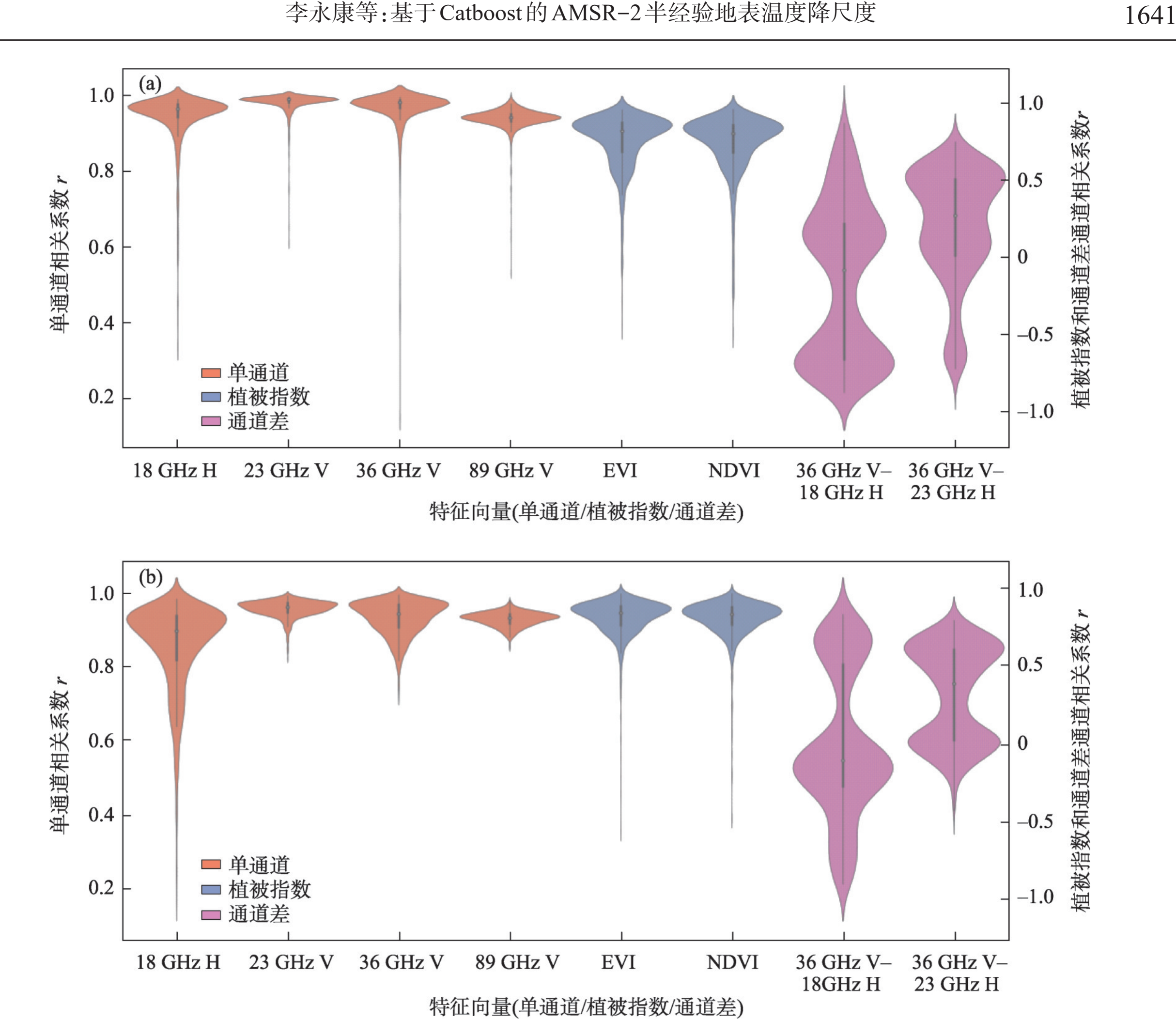


图3　日(a)、夜(b)间各特征因子相关性的频率统计

Fig. 3　Day (a) and night (b) correlations histogram of each factor

传输方程表示如下：

$$B_f(T_f)=\tau_f(\theta)\varepsilon_f B_f(T_s)+\left[1-\tau_f(\theta)\right](1-\tau_f(\theta)B_f\left(T_a^{\downarrow}\right)+\left[1-\tau_i(\theta)\right]B_f\left(T_a^{\uparrow}\right) \quad (1)$$

式中：T_s代表地表温度；T_a代表平均大气温度；T_f代表f频率的亮度温度；$\tau_f(\theta)$代表f频率在观测角θ的透射率；ε_f代表f频率的地表发射率；$B_f(T_s)$代表地表辐射；$T_a^{\downarrow}$与$T_a^{\uparrow}$分别为大气下行与上行辐射。

基于多通道辐射传输方程反演地表温度，每加入1个通道，将加入1个新的未知数，N个通道加入时，有N+1个未知数，因此基于被动微波的物理模型的地表温度反演为病态反演[14,32]。为解决此问题，Fily等学者假设同一频率下垂直极化发射率ε_V与水平极化发射率ε_H可简化为线性关系$\varepsilon_V=a\varepsilon_H+b$表示[15-17,33]，式(1)亦可简化为：

$$LST=\left[T_{b_V}-aT_{b_H}-(1-b-a)\tau_f(\theta)T_a^{\downarrow}\right]/\left[b\tau_f(\theta)\right] \quad (2)$$

式中：a、b为线性回归系数；T_b为微波亮温值；T_b的下标V、H分别代表垂直极化与水平极化。但是这种假设简化了先决条件，如恒定的大气校正、地表植被的吸收、忽略大气散射效应等[15-17,33]。

1.3.2　Catboost——被动微波地表温度反演的机器学习模型　通过机器学习方法多通道微波亮温反演地表温度，避免了简化垂直与水平亮温的相互关系，较好地解决病态反演问题。Catboost是在XG-boost、LightBMG等基础上新提出的一种GBDT算法[34]，能够较好的解决复杂的数据异构、噪声高、特征相互依赖等问题[24]。Catboost模型中特征向量提供信息高于目标向量[34]，而传统的GBDT算法通过平均目标向量来作为节点分割的准则[35]，因此Catboost能够利用到特征之间的联系，丰富特征维度，

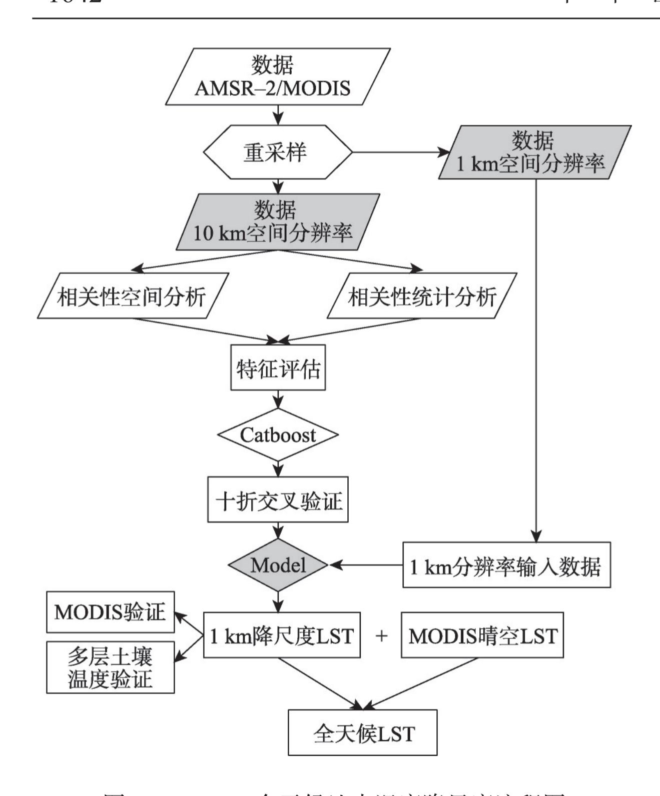


图4 AMSR-2全天候地表温度降尺度流程图

Fig. 4 Flow chart of the method for AMSR-2 all-weather land surface temperature downscaling

降低低频目标向量对结果影响;其为一种贪婪的机器学习方式,结合所有特征及所有特征组合建立决策树[24];此外,其通过排序梯度提升,添加先验分布项,去除训练集中噪声数据,克服传统GBDT的过拟合问题,解决预测梯度偏移[34]。

在模型实际运行中,贪婪的学习方式导致每迭代一次,都对目标变量进行统计分组,以及密集的数值特征将导致建立决策树的高计算负担。但以上问题在python接口Catboost算法通过GPU克服,且该算法支持多个GPU线程工作,分布式树学习实现对数据或特征并行处理。AMSR-2被动微波亮温具有多频率、双极化特征,Catboost算法能够在保证提取多通道特征的同时提高学习效率。

1.3.3 验证方法 本研究基于Catboost模型,对4通道组合向量进行训练,构建被动微波亮温、植被指数与MODIS地表温度之间的映射关系。Catboost算法的4通道模型初步评估,选用"十折交叉验证"[30-31]方案。十折交叉验证之初在统计学中用来测试算法模型的准确性。将古尔班通古特沙漠的训练数据随机分为10组,循环地将其中9组作为训练数据拟合模型,其余的1组作为验证。将10次的拟合精度均值作为本模型的最终精度。采用十折交叉验证能够有效降低系统误差、随机误差和粗差,使模型测试的精密度达到理论上的最高[36-37]。

特征向量的空间相关性统计分析。在像元尺度进行皮尔逊相关分析,获取各个特征向量的相关系数r及概率P,绘制相关系数r空间分布图,分析各特征因子与地表温度相关性的统计特征及空间特征。将决定系数R^2(Coefficient of determination)、平均绝对误差MAE(Mean absolute error)和均方根误差RMSE作为Catboost模型评价指标,进行超参数搜索、验证及对比。

模型的构建与评价分为3步:(1)以MYD11A1作为目标向量的Catboost模型构建,通过其Python接口输出Catboost 4通道模型的十折交叉验证,对模型进行初步评价;(2)根据Catboost 4通道模型,将降尺度后的1 km地表温度数据与MYD11A1晴空像元对比分析、评价精度;(3)将降尺度地温与站点6层地温分析,确定各层土温与被动微波降尺度地表温度相关性。

2 结果分析

2.1 降尺度特征向量空间相关性统计分析

从相关性的空间分布(图2)及相关性的频率统计(图3)两方面对特征因子与地表温度的相互关系分析表明:

(1)古尔班通古特沙漠日、夜间单通道微波与地表温度的相关性呈现绿洲、沙漠分异显著,主要表现为沙漠内部相关性较高,周边绿洲较低;与夜间相关性空间特征相比,日间的绿洲与沙漠分异性更强。

(2)EVI与NDVI相关性空间特征相似,能较好描述沙漠与周边绿洲分异性(沙漠区>绿洲区),但描述能力低于单通道亮温;相关性均值EVI>NDVI(日间EVI与NDVI分别为0.76、0.74,夜间分别为0.808、0.798),因此在模型构建中选用EVI作为模型构建的辅助数据。

(3)区域各特征因子相关性均值日间:23.8 GHz V(0.98)>36.5 GHz V(0.95)>18.7 GHz H(0.94)>89 GHz V(0.93),36.5-23.8 GHz V(0.31)>36.5 V-18.7 GHz H(-0.02),EVI(0.76)>NDVI(0.74);夜间:23.8 GHz V(0.96)>36.5 GHz(0.94)>89 GHz(0.93)>18.7 GHz(0.86),36.5-23.8 GHz V

(0.20)>36.5 V-18.7 GHz H(-0.16),EVI(0.808)>NDVI(0.798)。

36.5 V-18.7 H的通道组合与地表温度相关性空间特征日夜间差异较大,且分布完全相反。这可能是由日夜间沙漠区地表水分差异引起,沙漠日夜间地表温度差异较大,日间迅速增温,达到较高温度,沙漠化严重的西部水分蒸发缺失严重,而有稀疏植被覆盖的东部水分蒸发较少,因此相关性空间特征东西差异较大。夜间低温促进空气水分向地表水分运动,沙漠化严重的西部降温较快,有稀疏植被覆盖的东部较慢,因此卫星过境时(北京时间4:00)西部地表水分高于东部。此结果印证36.5 V-18.7 H与地表水分具有较高的相关性[20,22]。36.5 V-23.8 H在周边绿洲与地表温度呈负相关,沙漠内部呈正相关,其相关性空间特征分异性显著,这是由于36.5 V-23.8 H与大气水汽极度相关引起的[20,22,38]。

在像元尺度对相关系数r、概率P进行了计算,除古尔班通古特沙漠外西北方向盐场聚集区,单通道概率P均呈极显著水平,表明盐场会严重影响被动微波与地表温度的映射关系。

2.2 Catboost 4通道模型十折交叉验证

十折交叉验证结果(图5)表明,基于Catboost的4通道模型能够较好地构建10 km尺度下被动微波亮温与地表温度之间的映射关系,且表现出鲁棒性。日间R^2、MAE、RMSE均值分别为0.973、2.95 K、3.80 K,夜间R^2、MAE、RMSE均值分别为0.964、2.24 K、3.04 K,表明日夜间4通道Catboost模型精度均较高,泛化能力在10 km尺度下得到了较好的验证,且夜间交叉验证结果优于日间。日间最大值与最小值差分别为0.001、0.07 K、0.13 K;夜间最大值与最小值差分别为0.001、0.04 K、0.06 K,日夜间均较小,表明模型具有较高的稳定性。

2.3 降尺度精度验证及特征因子重要性评估

2.3.1 精度验证 MODIS为像元尺度(1 km×1 km)的降尺度地表温度验证提供了真值数据,Wang等[39]获得了2000—2001年的地表真值数据,对MODIS数据进行了评估,结果表明MODIS的地表温度产品精度在1 K以内;于文凭等[40]以黑河流域为研究区对MODIS地表温度产品进行精度验证,表明平均绝对误差小于2.2 K,可以用来间接评估反演产品的质量。因此,可选取MODIS质量average LST error ≤1 K的晴空像元对反演的地表温度验证。

通过MODIS质量average LST error≤1 K的晴空像元,对地表温度降尺度结果进行精度验证(图6),日间与夜间R^2分别为0.977、0.980,RMSE分别为3.69 K、2.38 K,MAE分别为2.71 K、1.70 K。表明基于4通道亮温的Catboost降尺度结果具有较高精度,可满足地表温度基本应用需求。

夜间地表温度降尺度结果精度高于日间。原因可能是古尔班通古特沙漠日间升温较快,日间过境时间(约15:00)温度升至较高水平,大气、土壤水分随温度升高降至较低水平,因此,4通道模型消除

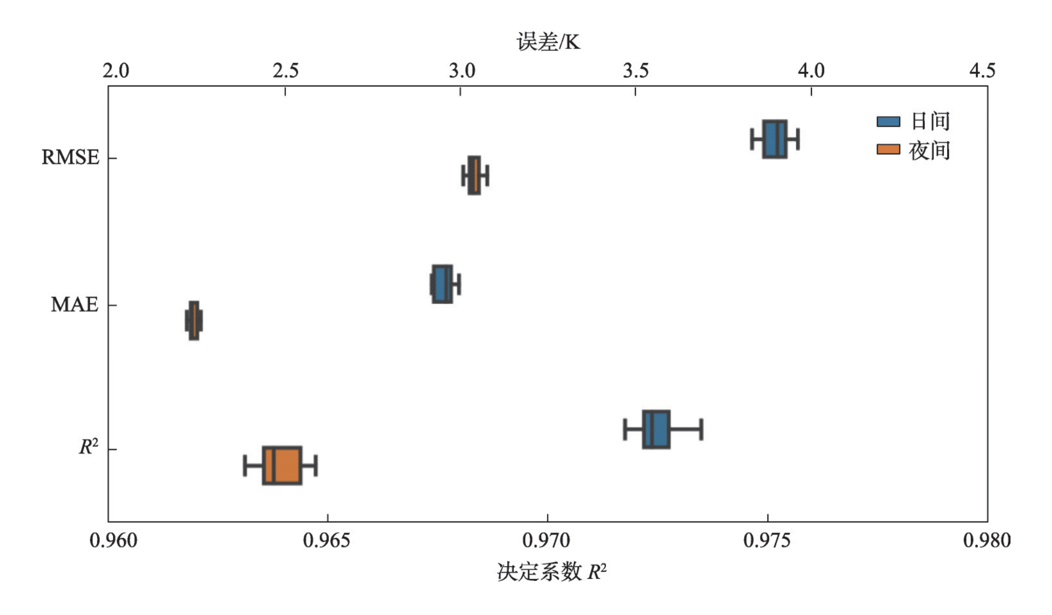


图5 日间、夜间Catboost十折交叉验证结果

Fig. 5 The 10-fold cross-validation result of Catboost in day and night

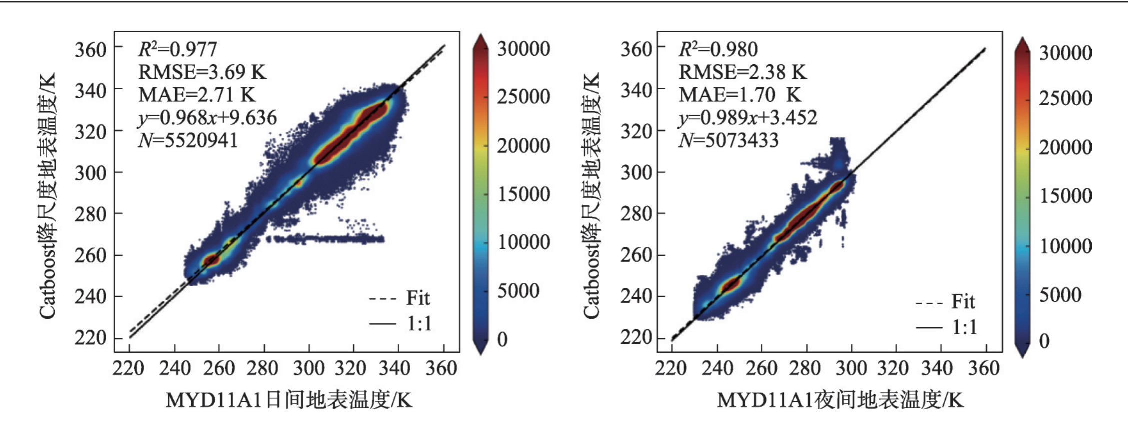


图6 基于日间、夜间Catboost半经验模型的地表温度降尺度结果与MYD11A1散点图

Fig. 6 The scatter plots for downscaled LST derived from day/night Catboost model with MYD11A1

大气、地表水分干扰的优势在日间被消弱;夜间过境时间(约4:00)沙漠温度降低至较低水平,大气、地表水分随温度降低升至较高水平,4通道模型对水汽的校正效应能够提高LST反演精度。此结果与36.5 V-18.7 H日夜间相关性分析一致。

2.3.2 重要性分析 由图7可知,日间特征向量重要性表现为:89 GHz V(37.7) > 36 GHz V(23.8) > 36 GHz V-23 GHz V(15.6) > EVI(15.4) > 36-23 GHz V(7.5);夜间特征向量重要性表现为:89 GHz V(47.2) > 36-23 GHz V(19.2) > 36 GHz V(13.3) > EVI(12.6) > 36-23 GHz V(7.8)。

单通道相关性统计结果与Catboost重要性分析结果存在差异。日间、夜间36 GHz垂直极化地表相关性均高于89 GHz,但重要性分析中,89 GHz高于36 GHz约1.5~3倍。日间36 GHz与地表温度的相关性均值(0.95)高于夜间相关性均值(0.94),而相较于日间重要性,夜间降低了约44%;日间89 GHz与地表温度的相关性均值(0.93)与夜间相等,而相较于日间重要性,夜间89 GHz占比增加约25%。表明相关性较低的单通道可能为多通道反演地表温度提供更多信息,且单微波通道与地表温度的相关性高低不能作为Catboost模型反演地表温度特征因子选择的依据。

2.4 降尺度结果的6层地温相关性分析

由于不同微波通道的穿透性具有较大差异,因此多通道亮温反演地表温度数据,往往缺乏对反演结果深度的定量描述,因此选取阜康荒漠生态站6层地温数据与对应位置的降尺度结果进行相关性分析(图8)。日夜间降尺度结果与6层地温的$P<0.001$,呈极显著相关;日间各层地温的相关性(r)5 cm(0.980) > 10 cm(0.977) > 15 cm(0.971) > 20 cm(0.969) > 40 cm(0.957) > 100 cm(0.899),夜间各层地温的相关性(r)5 cm(0.981) = 10 cm(0.981) = 15 cm(0.981) > 20 cm(0.980) > 40 cm(0.975) > 100 cm(0.928),相关性r均在0.8以上,呈极强相关。表明日夜间各深度的土壤温度,均与AMSR-2降尺度结果具有较高的相关性,且随着深度的增加,均方根误差呈增长趋势。

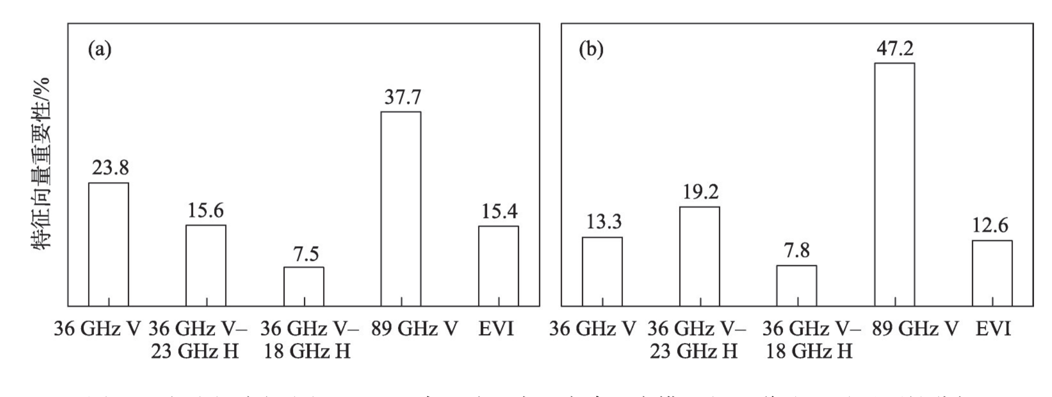


图7 日间(a)、夜间(b) Catboost半经验地表温度降尺度模型(4通道及EVI)重要性分析

Fig. 7 The importance bar of model (Catboost with 4 channels and EVI) feature vectors respectively in day (a) and night (b)

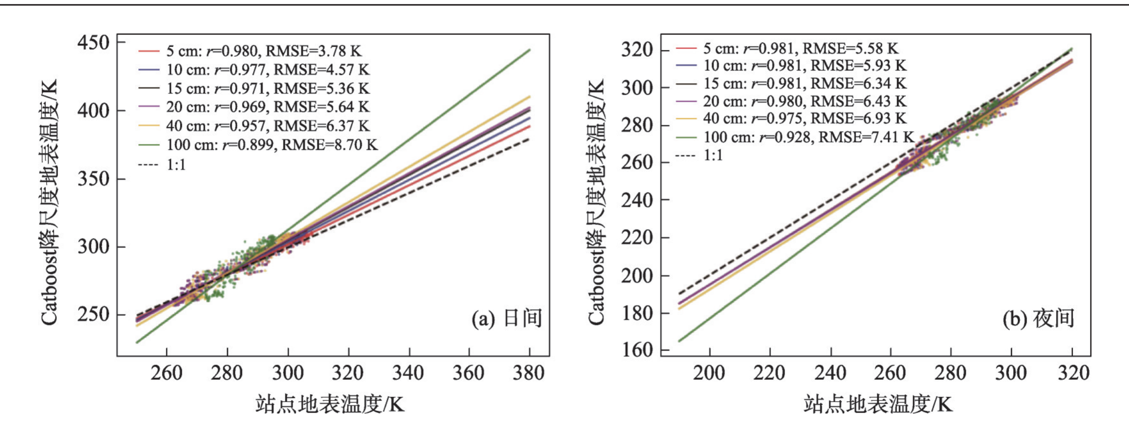


图8　降尺度结果与6层地温相关性分析

Fig. 8　Correlation plot of downscaled temperature based on 6 layers station data

日夜间反演结果与不同深度相关性变化差异主要存在于5、10、15、20 cm的4层。日间卫星过境时(约15:00),太阳辐射能量使土壤各层温度分异显著,随深度增加逐层降低特征;夜间过境时(约4:00),土壤缺乏太阳辐射能量输入,各层土壤温度均较低且分异性弱。多层地温的相关性检验结果显示,降尺度结果与各层土壤温度均具有较高的相关性。此结论可用于被动微波亮温反演多层土壤温度提供理论依据,结合本研究成果,可以避免热红外遥感的云强干扰,获取全天候的每天昼夜2次观测的6层温度产品。

2.5　基于Catboost的昼夜全天候地表温度产品

通过融合AMSR-2亮温反演得到日尺度的地表温度数据,填补MODIS由云导致的地表温度空缺像元,获得地表温度产品(图9)。当晴空时,使用MODIS提供的LST像元值,当有云雨出现时,使用AMSR-2微波亮温的反演地表温度。以8月13日与8月14日连续2 d为例进行展示,通过数据融合在昼夜尺度上获取了时间、空间连续的地表温度产品。

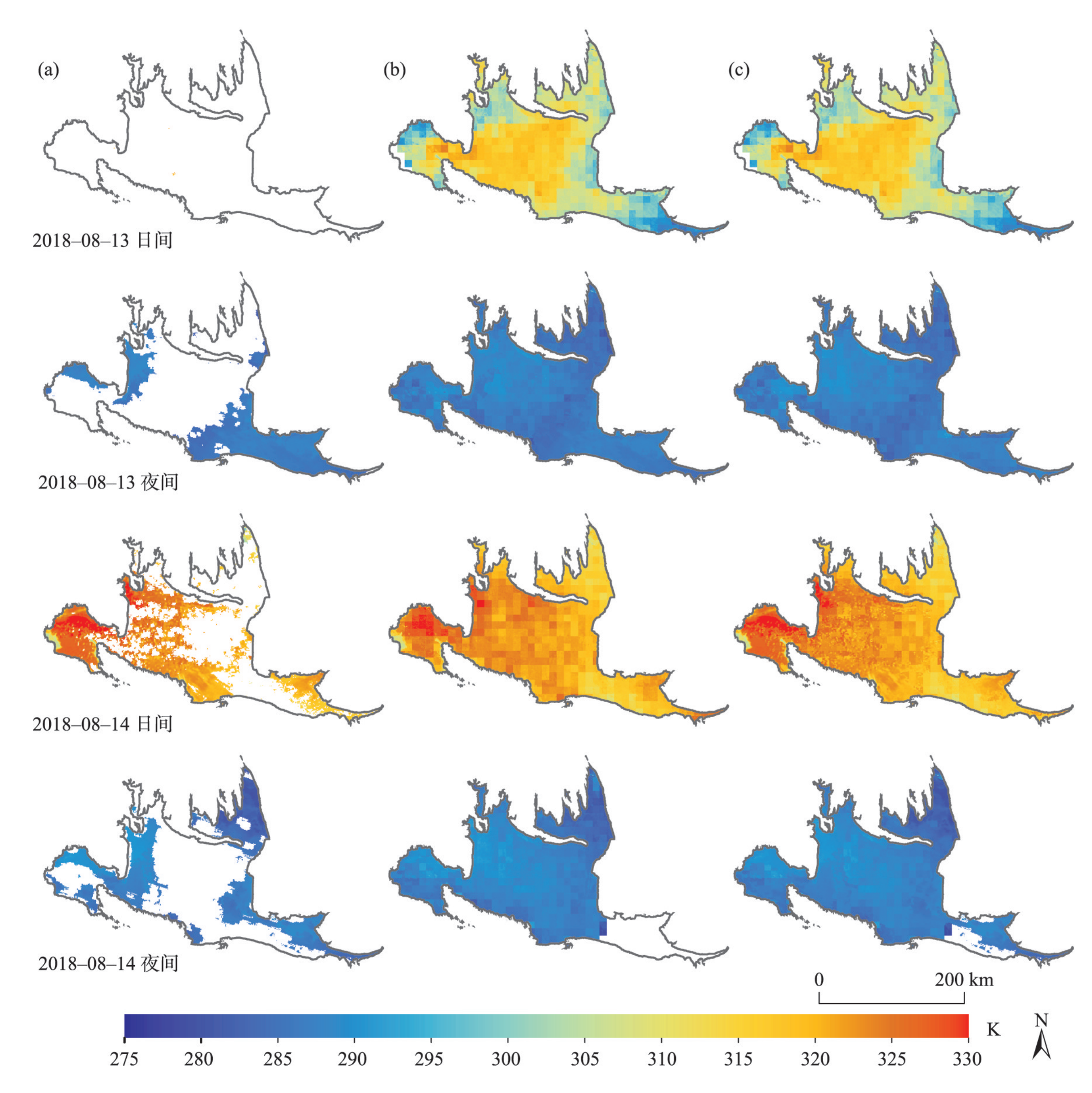
3　讨 论

比较Tan等[22]基于卷积神经网络垂直极化微波地表温度反演精度,本研究的R^2提高了0.08,RMSE降低了0.39 K,但是较全波段反演精度较低,且本文增加了夜间地表温度的反演。较Sun等[21]的4通道回归树的算法,日间、夜间R^2分别增加了0.075、0.076,RMSE分别减少了0.14 K、0.55 K,MAE分别减少了0.3 K、0.43 K;与该文章均表明夜间的精度高于日间,且对比证明Catboost对于被动微波与地表温度的拟合关系优于回归树;且在被动微波亮温反演地表温度的基础上进行了降尺度研究,构建了云下地表温度像元。此外,吸纳Huang等[41]的研究想法,将遥感地表温度与站点多层土壤温度建立相关关系,反演多层地温。通过被动微波亮温反演的地表温度产品与多层站点土壤温度相关性R^2更高(5 cm处达到0.98),但同时RMSE也更大。推测原因在于6.9 GHz最大干沙穿透深度达25 cm,能够更好的获取地下辐射相关信息,但训练的目标向量(MYD11A1)本身为表层温度,缺乏深层土壤温度数据来建立映射关系。

4　结 论

(1) 古尔班通古特沙漠日夜间各特征向量与地表温度的相关性空间分异明显,沙漠相关性高,绿洲相关性低,且日间分异性更强;盐矿覆盖会降低微波与地表温度相关性。区域各特征因子相关性均值表现为日间:23.8 GHz V(0.98)>36.5 GHz V(0.95)>18.7 GHz H(0.94)>89 GHz V(0.93),36.5 GHz V-23.8 GHz V(0.31)>36.5 V-18.7 GHz H(-0.02),EVI(0.76)>NDVI(0.74);夜间:23.8 GHz V(0.96)>36.5 GHz(0.94)>89 GHz(0.93)>18.7 GHz(0.86),36.5 GHz V-23.8 GHz V(0.20)>36.5 V-18.7 GHz H(-0.16),EVI(0.80)>NDVI(0.79)。

(2) 十折交叉验证表明Catboost 4通道模型建立的被动微波亮温与地表温度间映射关系表现出鲁棒性。日间R^2(coefficient of determination)、MAE(Mean Absolute Error)、RMSE(Root Mean Square Error)均值分别为0.973、2.95 K、3.80 K,夜间R^2、MAE、



注:(a)代表1 km分辨率晴空下MODIS LST(K);(b)代表1 km分辨率AMSR-2反演LST(K);(c)代表1 km分辨率MODIS与AMSR-2融合LST(K)。

图9 全天候地表温度产品合成图

Fig. 9 The integrated all-weather LST

RMSE均值分别为0.964、2.24 K、3.04 K,表明模型精度均较高。日间最大值与最小值差分别为0.001、0.07 K、0.13 K;夜间最大值与最小值差分别为0.001、0.04 K、0.06 K,日夜间均较小,表明模型具有较高的稳定性。

(3) 地表温度降尺度结果具有较高精度,MODIS验证日夜间 R^2 分别为0.977、0.980,RMSE分别为3.69 K、2.38 K,MAE分别为2.71 K、1.70 K。降尺度结果与站点7层地表温度进行了验证分析,结果表明与站点6层土壤温度均呈极显著正相关,并随着深度的增加相关系数 r 总体呈下降趋势,RMSE呈增加趋势,此结果为可微波获取全天候日、夜间多层土壤温度提供数值参考。

(4) 单通道相关性统计结果与重要性分析结果存在差异,日间、夜间36 GHz垂直极化地表相关性均高于89 GHz,89 GHz重要性高于36 GHz约1.5~3倍。日间36 GHz与地表温度的相关性均值(0.95)高于夜间相关性均值(0.94),而相较于日间重要性,夜间降低了约44%;日间89GHz与地表温度的相关性均值(0.93)与夜间相等,而相较于日间重要性,夜间89 GHz占比增加约25%。表明特征因子相关性统计结果不能直接作为Catboost被动微波反演地表温度特征选择依据。此外,各因子的重要性分析对于机理逆向评估具有重要意义。

致 谢:衷心感谢中国科学院新疆生态与地理研究所阜康荒漠生态试验站提供数据及程磊的帮助;感谢中国科学院空天信息创新研究院赵静对论文修改中的帮助!

参考文献(References):

[1] Li Z L, Tang B H, Wu H, et al. Satellite-derived land surface temperature: Current status and perspectives[J]. Remote Sensing of Environment, 2013, 131: 14–37.

[2] Anderson M, Norman J, Kustas W, et al. A thermal-based remote sensing technique for routine mapping of land-surface carbon, water and energy fluxes from field to regional scales[J]. Remote Sensing of Environment, 2008, 112(12): 4227–4241.

[3] Kalma J D, McVicar T R, McCabe M F. Estimating land surface evaporation: A review of methods using remotely sensed surface temperature data[J]. Surveys in Geophysics, 2008, 29(4): 421–469.

[4] Wang K, Li Z, Cribb M. Estimation of evaporative fraction from a combination of day and night land surface temperatures and NDVI: A new method to determine the Priestley-Taylor parameter[J]. Remote Sensing of Environment, 2006, 102(3): 293–305.

[5] Faqe I G. Urban Land Use land cover changes and their effect on land surface temperature: Case study using dohuk city in the Kurdistan Region of Iraq[J]. Climate, 2017, 5(1): 13–31.

[6] Wan Z, Wang P, Li X. Using MODIS land surface temperature and normalized difference vegetation index products for monitoring drought in the southern Great Plains, USA[J]. International Journal of Remote Sensing, 2010, 25(1): 61–72.

[7] Hain C R, Anderson M C. Estimating morning change in land surface temperature from MODIS day/night observations: applications for surface energy balance modeling[J]. Geophysical Research Letters, 2017, 44(19): 9723–9733.

[8] Neteler M. Estimating daily land surface temperatures in mountainous environments by reconstructed MODIS LST data[J]. Remote Sensing, 2010, 2(1): 333–351.

[9] Wang F, Qin Z, Song C, et al. An improved mono-window algorithm for land surface temperature retrieval from Landsat 8 thermal infrared sensor data[J]. Remote Sensing, 2015, 7(4): 4268–4289.

[10] Duan S B, Li Z L, Wang C, et al. Land-surface temperature retrieval from Landsat 8 single-channel thermal infrared data in combination with NCEP reanalysis data and ASTER GED product[J]. International Journal of Remote Sensing, 2018, 40(5–6): 1763–1778.

[11] Mao K, Shi J, Tang H, et al. A neural network technique for separating land surface emissivity and temperature from ASTER Imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2008, 46(1): 200–208.

[12] 何文英, 陈洪滨, 李军. 陆面微波比辐射率的卫星观测研究进展[J]. 地球物理学报, 2020, 63(10): 3573–3584. [He Wenying, Chen Hongbin, Li Jun. Development of land surface microwave emissivity retrieval using satellite observations[J]. Chinese Journal of Geophysics, 2020, 63(10): 3573–3584.]

[13] 钱博, 陆其峰, 杨素英, 等. 卫星遥感微波地表发射率研究综述[J]. 地球物理学进展, 2016, 31(3): 960–964. [Qian Bo, Lu Qifeng, Yang Suying, et al. Review on microwave land surface emissivity by satellite remote sensing[J]. Progress in Geophysics, 2016, 31(3): 960–964.]

[14] 汪子豪, 秦其明, 孙元亨, 等. 基于BP神经网络的地表温度空间降尺度方法[J]. 遥感技术与应用, 2018, 33(5): 793–802. [Wang Zihao, Qin Qiming, Sun Yuanheng, et al. Downscaling of remotely sensed land surface temperature with the bp neural network[J]. Remote Sensing Technology and Application, 2018, 33(5): 793–802.]

[15] McFarland M J, Miller R L, Neale C M U. Land surface temperature derived from the SSM/I passive microwave brightness temperatures[J]. IEEE Transactions on Geoscience and Remote Sensing, 1990, 28(5): 839–845.

[16] Fily M. A simple retrieval method for land surface temperature and fraction of water surface determination from satellite microwave brightness temperatures in sub-arctic areas[J]. Remote Sensing of Environment, 2003, 85(3): 328–338.

[17] Gao H, Fu R, Dickinson R E, et al. A practical method for retrieving land surface temperature from AMSR-E over the Amazon Forest[J]. IEEE Transactions on Geoscience and Remote Sensing, 2008, 46(1): 193–199.

[18] Njoku E G, Li L. Retrieval of land surface parameters using passive microwave measurements at 6–18 GHz[J]. IEEE Transactions on Geoscience and Remote Sensing, 1999, 37(1): 79–93.

[19] Mao K, Shi J, Li Z, et al. A physics-based statistical algorithm for retrieving land surface temperature from AMSR-E passive microwave data[J]. Science in China Series D: Earth Sciences, 2007, 50 (7): 1115–1120.

[20] 毛克彪, 王道龙, 李滋睿, 等. 利用AMSR–E被动微波数据反演地表温度的神经网络算法[J]. 高技术通讯, 2009, 19(11): 1195–1200. [Mao Kebiao, Wang Daolong, Li Zirui, et al. A neural network method for retrieving land-surface temperature from AMSR–E data[J]. Chinese High Technology Letters, 2009, 19(11): 1195–1200.]

[21] Sun D, Li Y, Zhan X, et al. Land surface temperature derivation under all sky conditions through integrating AMSR-E/AMSR-2 and MODIS/GOES observations[J]. Remote Sensing, 2019, 11(14): 1704.

[22] Tan J, Esmaeel N, Mao K, et al. Deep learning convolutional neural network for the retrieval of land surface temperature from AMSR2 data in China[J]. Sensors, 2019, 19(1): 2987–3007.

[23] Huang G, Wu L, Ma X, et al. Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions[J]. Journal of Hydrology, 2019, 574: 1029–1041.

[24] Zhang Y, Zhao Z, Zheng J. CatBoost: A new approach for estimat-

ing daily reference crop evapotranspiration in arid and semi-arid regions of Northern China[J]. Journal of Hydrology, 2020, 588 (125): 87–99.

[25] 姜琦刚, 杨秀艳, 杨长保, 等. 基于CatBoost算法的面向对象土地利用分类[J]. 吉林大学学报(信息科学版), 2020, 38(2): 185–191. [Jiang Qigang, Yang Xiuyan, Yang Changbao, et al. Object-oriented land use classification based on CatBoost algorithm[J]. Journal of Jilin University(Information Science Edition), 2020, 38 (2): 185–191.]

[26] 王新军, 赵成义, 杨瑞红, 等. 古尔班通古特沙漠南缘荒漠化过程演变的景观格局特征分析[J]. 干旱区地理, 2015, 38(6): 1213–1225. [Wang Xinjun, Zhao Chengyi, Yang Ruihong, et al. Landscape pattern characteristics of desertification evolution in southern Gurbantunggut Desert[J]. Arid Land Geography, 2015, 38 (6): 1213–1225.]

[27] Mamtimin A, Wang Y, Sayit H, et al. Seasonal variations of the near-surface atmospheric boundary layer structure in China's Gurbantünggüt Desert[J]. Advances in Meteorology, 2020: 6137237.

[28] 段呈, 吴玲, 王绍明, 等. 近30年古尔班通古特沙漠短命植物的时空格局[J]. 生态学报, 2017, 37(8): 2642–2652. [Duan Cheng, Wu Ling, Wang Shaoming, et al. Analysis of spatio-temporal patterns of ephemeral plants in the Gurbantünggüt Desert over the last 30 years[J]. Acta Ecologica Sinica, 2017, 37(8): 2642–2652.]

[29] 陈林, 王式功, 王莉莉. 新疆阜康地区秋季大气NO_x和O_3变化特征及影响要素[J]. 干旱气象, 2012, 30(3): 345–352. [Chen Lin, Wang Shigong, Wang Lili. Variation characteristics and influencing factors of NO_x and Ozone in autumn in Fukang Region of Xinjiang[J]. Journal of Arid Meteorology, 2012, 30(3): 345–352.]

[30] 范永东. 模型选择中的交叉验证方法综述[D]. 太原: 山西大学, 2013. [Fan Yongdong. A Summary of Cross-Validation in Model Selection[D]. Taiyuan: Shanxi University, 2013.]

[31] 郝静, 孙成, 郭兴宇, 等. 京津冀内陆平原区$PM_{2.5}$浓度时空变化定量模拟[J]. 环境科学, 2018, 39(4): 1455–1465. [Hao Jing, Sun Cheng, Guo Xingyu, et al. Simulation of the spatio-temporally resolved $PM_{2.5}$ aerosol mass concentration over the inland plain of the Beijing-Tianjin-Hebei region[J]. Environmental Science, 2018, 39(4): 1455–1465.]

[32] 金旭峰, 吴莹, 钱博. 基于AMSR2数据的青藏高原地表温度反演[J]. 地球物理学进展, 2020, 35(4): 1269–1275. [Jin Xufeng, Wu Ying, Qian Bo. Retrieval of land surface temperature from AMSR2 data over the Qinghai-Tibetan Plateau[J]. Progress in Geophysics, 2020, 35(4): 1269–1275.]

[33] Bhagat V S. Space-borne passive microwave remote sensing of soil moisture: A review[J]. Recent Progress in Space Technology, 2015, 4(2): 119–150.

[34] Prokhorenkova L, Gusev G, Vorobev A, et al. CatBoost: Unbiased boosting with categorical features[J]. Advances in Neural Signal Processing, 2018: 6638–6648.

[35] Zhang K, Scholkopf B, Muandet K, et al. Domain adaptation under target and conditional shift[C]//30th International Conference on Machine Learning, ICML, 2013: 1856–1864.

[36] Colliander A, Jackson T J, Bindlish R, et al. Validation of SMAP surface soil moisture products with core validation sites[J]. Remote Sensing of Environment, 2017, 191: 215–231.

[37] 陶本藻, 邱卫宁, 张书毕, 等. 误差理论与测量平差[M]. 武汉: 武汉大学出版社, 2012. [Tao Benzao, Qiu Weining, Zhang Shubi, et al. Error Theory and Measurement Adjustment[M]. Wuhan: Wuhan University Press, 2012.]

[38] 毛克彪, 胡德勇, 黄健熙, 等. 针对被动微波AMSR-E数据的土壤水分反演算法[J]. 高技术通讯, 2010, 20(6): 651–659. [Mao Kebiao, Hu Deyong, Huang Jianxi, et al. A algorithm for retrieving soil moisture from AMSR-E passive microwave data[J]. Chinese High Technology Letters, 2010, 20(6): 651–659.]

[39] Wang W, Liang S, Meyers T. Validating MODIS land surface temperature products using long-term nighttime ground measurements [J]. Remote Sensing of Environment, 2008, 112(3): 623–635.

[40] 于文凭, 马明国. MODIS地表温度产品的验证研究——以黑河流域为例[J]. 遥感技术与应用, 2011, 26(6): 705–712. [Yu Wenping, Ma Mingguo. Validation of the MODIS land surface temperature products: A case study of the Heihe River basin[J]. Remote Sensing Technology and Application, 2011, 26(6): 705–712.]

[41] Huang R, Huang J-x, Zhang C, et al. Soil temperature estimation at different depths, using remotely-sensed data[J]. Journal of Integrative Agriculture, 2020, 19(1): 277–290.

Downscaling land surface temperature through AMSR−2 passive microwave observations by Catboost semiempirical algorithms

LI Yongkang[1,2], WANG Xinjun[1], MA Yanfei[3], HU Guifeng[1], GUI Haiyue[1], ZHANG Guanhong[1]

(1. College of Grassland and Environment Sciences of Xinjiang Agricultural University, Xinjiang Key Laboratory of Soil and Plant Ecological Processes, Urumqi 830052, Xinjiang, China; 2. Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; 3. Department of Geography of Handan University, Handan 056005, Hebei, China)

Abstract: This study aimed to fill in the missing pixels of MYD11A1 in the Gurbantunggut Desert and provide a theoretical basis for obtaining all-weather and multilayer soil temperatures during daytime and nighttime. We explored the feasibility of the CatBoost algorithm for the spatial downscaling of passive microwave surface temperature by using the four-channel passive microwave brightness temperature and MODIS vegetation index of the 2019 AMSR−2. Results show that (1) the spatial differentiation between feature vectors and surface temperature is evident during daytime and nighttime in the Gurbantunggut Desert. It indicates high desert correlation, low oasis correlation, and strong daytime differentiation. The salt mine coverage reduces the correlation between passive microwave brightness temperature and surface temperature. (2) The mapping relation between the passive microwave brightness temperature and surface temperature from the four-channel CatBoost model is robust. The accuracy of the downscaled results is high, with daytime-nighttime R^2 of 0.977 and 0.980, RMSE of 3.69 and 2.38 K, and MAE of 2.71 K and 1.70 K, respectively. (3) Single-channel correlation and importance analysis results are different, suggesting that feature correlation results cannot be directly used as a basis for selecting surface temperature features with the CatBoost passive microwave inversion of LST. (4) Downscaling LST is significantly and positively correlated with the soil temperature in six layers from the Fukang site. The correlation coefficient decreases, and RMSE increases as depth increases.

Keywords: land surface temperature; AMSR−2; Catboost; downscale; Gurbantunggut Desert